Design and Analysis for Genetic Study of Complex Traits

Tutors
Jing Hua Zhao, MRC Epidemiology Unit, Cambridge, UK
Fuzhong Xue, Shandong University, Jinan, China

Synopsis

Owing to recent advances in genotyping and sequencing technologies and the successes of several international collaborative projects, there is a considerable interest in genetic analysis of complex traits which include common diseases and other quantitative measurements. The analysis customarily involves a large number of single-nucleotide polymorphisms (SNPs), the most abundant genetic variants in human genome.

This course intends to give an overview of approaches for genetic analysis of complex traits, including heritability, segregation, linkage and association studies, paying attention to the statistical models, some successful stories and indication of their limitations. A particular focus is on the study design and analysis of genome-wide association studies (GWAS), and the instructors' own involvement in such analysis will be described. A complementary part of this tutorial concerns about computer software in these analyses. Statistical and computational challenges are expected to be exposed through both parts.

Topics associated with and motivated from the case studies range from fundamental concepts such as measurement of risk and heritability to analysis of genomic data such as Hardy-Weinberg equilibrium, linkage equilibrium to more sophisticated modeling such as prospective and retrospective models of haplotypes, gene-gene, gene-environmental interactions and pathways. Related aspects include haplotype analysis, imputation of genotypes and meta-analysis, merits of some frequency-based and Bayesian methods as used in our GWAS of obesity.

The potential attendees will be researchers with basic knowledge in statistics and computing who wish to get involved with or improve their understanding of genetic data analysis. However, it will also be useful to professionals and researchers actively engaged in analysis of genetic data and/or development of computational tools in R or other environments. It is expected that course materials will refresh and interact with attendees' views on design and analysis of genomic data in humans while generating interest to researchers in plant and animal sciences.

Course Schedule (Mon 20 Oct. 2008 – Sun 26 Oct. 2008)

8:00 am – 12:00 noon (break at 9:45-10:15)

20 Oct. 2008	Overview references
21 Oct. 2008	Analysis of family resemblance and segregation references
22 Oct. 2008	Linkage analysis references
23 Oct. 2008	Association analysis references
24 Oct. 2008	Issues in association analysis references
25 Oct. 2008	Study design references
26 Oct. 2008	Advanced topics and summary references

1. Overview

Course outline
Basic terminology
Genetic epidemiology and study of complex traits
Genome projects and genomewide association studies (GWAS)
A primer of statistics and statistical computing, R, SAS, Stata

2. Analysis of family resemblance and segregation

Genetic relationships and gene identity
Path analysis of twin and nuclear family data
Commingling analysis
Parametric models for family data
Other models for family data: test for familial aggregation, GEE, regressive and mixed models
Case studies, gene expression levels and aging in CEPH families
Practice: PATHMIX/Mx/Mplus, kinship/multic, SAGE/SOLAR

3. Linkage analysis

Scope and concepts
Parametric methods
Nonparametric methods
Issues of linkage analysis
Case studies
Practice: LINKAGE, GENEHUNTER, Merlin, SAGE/SOLAR

4. Association analysis

Scope and concepts
Population-based association analysis
Family-based association analysis
Case studies
Practice: ETDT, EHPLUS, PLINK, HaploView, R, SAS, Stata

5. Issues in association analysis

Handling of large data
Population stratification
Multiple testing
Meta-analysis

Gene characterization
Practice: R, SAS and Stata

6. Study design

A comparison of linkage and association designs
Case-cohort design
Practice: SLINK, QUANTO, R

7. Advanced topics and summary

Other methods for handling high dimensional data
Exploratory and confirmatory analysis of pathways
Practice: Mplus, R

Books for probability & statistics

Agresti A. Categorical Data Analysis. 2nd Edition. Wiley 2002

Armitage P, T Colton (Eds). Encylcopedia of Biostatistics, 2nd Edition, Wiley 2005

Casella G, RL Berger. Statistical Inference. 2nd Edition. Duxbury 2002

Congdon P. Applied Bayesian Modelling. Wiley 2003

Diggle PJ, P Heagerty, K-Y Liang, SL Zeger. Analysis of Longitudinal Data, 2nd Edition. Oxford University Press 2002

Gelman A, JB Carlin, HS Stern, DB Rubin. Bayesian Data Analysis. 2nd Edition, Chapman & Hall/CRC 2004

Hosmer DW, S Lemeshow. Applied Logistic Regression, 2nd Edition. Wiley 2000

Lachin JM. Biostatistical Methods. Wiley 2000

McCulloch CE, Searle SR. Generalized, Linear, and Mixed Models. Wiley 2001

Paolella MS. Intermediate Probability: A Computational Approach. Wiley 2007

Ross SM. Introduction to Probability Models, 9th Edition. Elsevier 2007

Schafer JL. Analysis of Incomplete Multivariate Data. CRC Press 1997

Singer JD, JB Willett. Applied Longitudinal Data Analysis: Modelling Change and Event Occurrene. Oxford University Press 2003

Snijders T, R Bosker. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. SAGE 1999

Weisberg S. Applied Linear Regression, 3rd Edition. Wiley 2005

General

Bodmer W, C Bonilla. Common and rare variants in multifactorial susceptibility to common diseases. Nature Genetics 40:695-701, 2008

Burton PR, MD Tobin MD, JL Hopper. Key concepts in genetic epidemiology. Lancet 366: 941-51, 2005

Davey Smith G, S Ebrahim, S Lewis, AL Hansell, LJ Palmer, PR Burton. Genetic epidemiology and public health: hope, hype, and future prospects. Lancet 366: 1484-98, 2005

Elston RC, M Anne Spence. Advances in statistical human genetics over the last 25 years. Statistics in Medicine, 25:3049-3080, 2006

Falconer DS, TFK Mackay. Introduction to Quantitative Genetics, 4th Edition. Longman 1996

Gianola D, Fernando RL, Stella A: Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173:1761-1776. 2006

Hartl DL, AG Clark. Principles of Population Genetics. Sinauer 1997

Horton NJ, SR Lipsitz. Multiple imputation in practice: comparison of software packages for regression models with missing variables. American Statistician 55: 244-54, 2001

Khoury M, T Beaty, B Cohen. Fundamentals of Genetic Epidemiology. Oxford University Press 1993

King RC, WD Stansfield, PK Mulligan. A Dictionary of Genetics, 7th Edition. Oxford University Press 2006

Lander ES, NJ Schork. Genetic dissection of complex traits. Science 265:2037-2048, 1994

Lockhart DJ, EA Winzeler. Genomics, gene expression and DNA arrays. Nature 405: 827-36, 2000

Manolio TA, LD Brooks, FS Collins. A HapMap harvest of insights into the genetics of common disease. Journal of Clinical Investigation 118:1590-1605, 2008

Morton NE. Outline of Genetic Epidemiology. Karger 1982

Morton NE, DC Rao, J-M Lalouel. Methods in Genetic Epidemiology, Karger 1983

Murray RK, DK Granner, PA Mayes, VW Rodwell. Harper's Illustrated Biochemistry, 26th Edition. McGraw-Hill 2003

Patchter L, B Sturmfels (Eds). Algebraic Statistics for Computational Biology. Cambridge University Press 2005

Sham PC. Genetic epidemiology. Br Med Bull 52:408-433, 1996

Sham PC. Statistics in Human Genetics. Edward Arnold 1998

Thomas DC. Statistical Methods in Genetic Epidemiology. Oxford, Oxford University Press 2004

Wacholder S, S Chanock, M Garcia-Closas, L El ghomli, N Rothman: Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. Journal of the National Cancer Institute 96:434-442, 2004

Wahba G: Soft and hard classification by reproducing kernel Hilbert space methods. Proceedings of the National Academy of Science USA 99:16524-16530, 2002

Zhao JH, Q Tan. Genetic dissection of complex traits *in silico*: approaches, problems and solutions. Current Bioinformatics 1:359-369, 2006

Association

Altshuler D, LD Brooks, A Chakravarti, FS Collins, MJ Daly, P Donnelly. A haplotype map of the human genome. Nature 437: 1299-320, 2005

Aulchenko YS, DJ de Koning, C Haley. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 177:577-585, 2007

Balding DJ. A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7:781-791, 2006

Barnes C, V Plagnol, T Fitzgerald, R Redon, J Marchini, D Clayton, ME Hurles. A robust statistical method for case-control association testing with copy number variation 40:1245-1252, 2008

Barrett JC, B Fry, J Maller, MJ Daly. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265, 2005

Benjamini Y, Y Hochberg. Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B 57: 289-300, 1995

Chanock SJ, T Manolio, M Boehnke, E Boerwinkle, DJ Hunter, G Thomas, JN Hirschhorn, G Abecasis, D Altshuler, JE Bailey-Wilson, et al. Replicating genotype-phenotype associations. Nature 447:655-660, 2007

Chen WM, GR Abecasis. Family-based association tests for genomewide association scans. American Journal of Human Genetics 81:913-926, 2007

Devlin B, K Roeder. Genomic control for association studies. Biometrics 55: 997-1004, 1999

Dudoit S, JP Shaffer, JC Boldrick. Multiple hypothesis testing in microarray experiments. Statistical Science 18:71-103, 2003

Efron B. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. Journal of the American Statistical Association 99:96-104, 2004

Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. Genetic Epidemiology 25:115-121,2003

Frayling TM, NJ Timpson, MN Weedon, E Zeggini, RM Freathy, CM Lindgren, JR Perry, KS Elliott, H Lango, NW Rayner, et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science, 316, 889-894

Hattersley AT, MI McCarthy. What makes a good genetic association study? Lancet 366:1315-23, 2005

Hopper JL, DT Bishop, DF Easton. Population-based family studies in genetic epidemiology. Lancet 366:1397-406, 2005

Klein RJ, C Zeiss, EY Chew, JY Tsai, RS Sackler, C Haynes, AK Henning, JP SanGiovanni, SM Mane, ST Mayne, MB Bracken, FL Ferris, J Ott, C Barnstable, J Hoh. Complement factor H polymorphism in age-related macular degeneration. Science 308: 385-89, 2005

Korn JM, FG Kuruvilla, SA McCarroll, A Wysoker, J Nemesh, S Cawley, E Hubbell, J Veitch, PJ Collins, K Darvishi, C Lee, MM Nizzari, SB Gabriel, S Purcell, MJ Daly, D Altshuler. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nature Genetics 40:1253-1260, 2008

Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genetics 22:139-144, 1999

Linsel-Nitschke P, A Gotz, J Erdmann, I Graenne, P Braund, C Hengstenberg, et al. Lifelong reduction of LDL-cholesterol related to a common variant in the LDL-receptor gene decreases the risk of coronary artery disease—a Mendelian randomisation study. PLoS One 3:e2986, 2008

Loos RJ, CM Lindgren, S Li, E Wheeler, JH Zhao, I Prokopenko, M Inouye, RM Freathy, AP Attwood, JS Beckmann, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nature Genetics 40:768-775, 2008

Marchini J, B Howie, S Myers, G McVean, P Donnelly. A new multipoint method for genome-wide association studies via imputation of genotypes. Nature Genetics 39:906-913, 2007

McCarthy MI, G Abecasis, LR Cardon, DB Goldstein, J Little, JPA Ioannidis, and JN Hirschhorn. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics 9:356-369, 2008

Palmer LJ, LR Cardon. Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet 366:1223-34, 2005

Pearson TA, TA Manolio. How to interpret a genome-wide association study. *JAMA* 299:1335-1344, 2008

Price AL, NJ Patterson, RM Plenge, ME Weinblatt, NA Shadick, D Reich. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904-909, 2006

Purcell S, B Neale, K Todd-Brown, L Thomas, MA Ferreira, D Bender, J Maller, P Sklar, PL de Bakker, MJ Daly, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American.Journal of.Human Genetics 81:559-575, 2007

Strachan T, AP Read. Human Molecular Genetics 2. BIOS 1999

Tabor HK, N Risch, RM Myers. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews Genetics 3:391-97, 2002

Tan Q, L Christiansen, C-B Andersen, JH Zhao, S Li, AK Torben, K Christensen. Retrospective analysis of main and interaction effects in genetic association studies of human complex traits. BMC Genetics 8:70, 2007

The International HapMap Project. Nature 426:789-96, 2003

The WTCCC. Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661-678, 2007

Wang J, J Cohen. Letter to the Editor. Annals of Human Genetics 63:409, 1999

Zhao H. Family-based association studies. Statistical Methods in Medical Research 9:563-987, 2000

Zhao JH, D Curtis, PC Sham. Model-free and permutation tests for allelic associations. Human Heredity 50: 133-139, 2000

Familial aggregation

Boyle CR, RC Elston. Multifactorial genetic models for quantitative traits in humans. Biometrics 35:55-68, 1979

Houwing-Duistermaat JJ, BHF Derkx, FR Rosendaal, HC van Houwelingen. Testing familial aggregation. Biometrics 51:1292-1301, 1995

Neale MC, LR Cardon. Methodology for genetic studies of twins and families. Kluwer 1992

Rao DC, NE Morton, S Yee. Analysis of family resemblance. II. A linear model for familial correlation. American Journal of Human Genetics 26:331-59, 1974

Self SG, G Longton, KJ Kopecky, K-Y Liang. On estimating HLA/disease association with application to a study of aplastic anemia. Biometrics 47: 55-61, 1991

Tan Q, JH Zhao, S Li, L Christiansen, TA Kruse, K Christensen. Differential and correlation analyses of microarray gene expression data in the CEPH Utah families. Genomics 92:94-100

Wardle J, S Carnell, CMA Haworth, R Plomin. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. American Journal of Clinical Nutrition 87:398-404, 2008

Commingling and segregation

Bonney GE. A note on the basis of regressive models for genetic analysis. Genetic Epidemiology. S1:37-42, 1986

Cannings C, EA Thompson, MH Skolnick. Probability functions on complex pedigrees. Advances in Applied Probability 10:26-61, 1978

Konigsberg LW, CM Kammerer, JW MacCluer. Segregation analysis of quantitative traits in nuclear families: comparison of three program packages. Genetic Epidemiology 6:713-726, 1989

MacLean CJ, NE Morton, RC Elston, S Yee. Skewness in commingled distributions. Biometrics 32:695-99, 1976

Morton N, C MacLean. Analysis of family resemblance. III. Complex segregation analysis of quantitative traits. American Journal of Human Genetics 27: 365-84, 1974

Linkage

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. *Merlin*—rapid analysis of dense genetic maps using sparse gene flow trees. Nature Genetics 30: 97-101, 2002

Almasy L, J Blangero. Multipoint quantitative-trait linkage analysis in general pedigrees. American Journal of Human Genetics 62:1198-211, 1998

Blangero J, HH Goring, JW Kent, JT Williams, CP Peterson, L Almasy, TD Dyer. Quantitative trait nucleotide analysis using Bayesian model selection. Human Biology 77:541-559, 2005

Chen WM, GR Abecasis GR. Estimating the power of variance component linkage analysis in large pedigrees. Genet Epidemiol 30:471-484, 2006

Dawn Teare M, JH Barrett. Genetic linkage studies. Lancet 366:1036-44, 2005

Dib C, S Faure, C Fizames, D Samson, N Drouot, A Vignal, P Millasseau, S Marc, J Hazan, E Seboun, et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152-154, 1996

Elston RC, J Stewart. A general model for the genetic analysis of pedigree data. Human Heredity 21:523-42, 1971

Gudbjartsson DF, K Jonasson, ML Frigge, A Kong. Allegro, a new computer program for multipoint linkage analysis. Nature Genetics 25: 12-13, 2000

Gusella JF, NS Wexler, PM Conneally, SL Naylor, MA Anderson, RE Tanzi, PC Watkins, K Ottina, MR Wallace, AV Sakaguchi, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234-238, 1983

Haseman JK, RC Elston. The investigation of linkage between a quantitative trait and a marker locus. Behavior Genetics 2:3-19, 1972

Heath SC. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. American Journal of Human Genetics 61:748-760, 1997

Kruglyak L, MJ Daly, MP Reeve-Daly, ES Lander. Parametric and Nonparametric Linkage Analysis: A Unified Multipoint Approach. American Journal of Human Genetics 58:1347-63, 1996

Lander ES, L Kruglyak. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 11:241-247, 1995

Lander ES, P Green. Construction of multilocus genetic maps in humans. Proceedings of the National Academy of Sciences of the USA 84:2363-67, 1987

Lathrop G, J Lalouel, C Julier, J Ott. Strategies for multilocus linkage analysis in humans. Proceedings of the National Academy of Science USA 81:3443-46, 1984

Morton NE. Sequential tests for the detection of linkage. American Journal of Human Genetics 7:277-318, 1955

Ott J. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. American Journal of Human Genetics 26:588-97, 1974

Ott J. Analysis of Human Genetic Linkage, 3rd Edition. Johns Hopkins University Press 1999

Shih M-C, Whittemore AS. *Allele-sharing among affected relatives: non-parametric methods for identifying genes.* Stat Meth Med Res. 10:27-55, 2001

Terwilliger J, J Ott. Handboook of Human Genetic Linkage. Johns Hopkins University Press 1994

Meta-analysis

Demidenko E. Mixed Models: Theory and Applications. Wiley 2004

DerSimonian R, N Laird. Meta-analysis in clinical trials. Control Clinical Trials **7**:177-188, 1986

Etzel CJ, R Gueera. Meta-analysis of genetic-linkage analysis of quantitative-trait loci. American Journal of Human Genetics 71:56-65. 2002

Gurevitch J, LV Hedges. Statistical issues in ecological meta-analyses. Ecology 80:1142-1149, 1999

Higgins JPT, A Whitehead, RM Turner, RZ Omar, SG Thompson. Meta-analysis of continuous outcome data from individual patients. Statistics in Medicine

20:2219–2241, 2001

Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. Statistics in Medicine. 18: 321-359, 1999

Stukel TA, E Demidenko, J Dykes, MR Karagas. Two-stage methods for the analysis of pooled data. Statistics in Medicine 20:2115-2130, 2001

van Houwelingen HC, LR Arends, T Stijnen. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine 21:589-624, 2002

Whitehead A. Meta-analysis of Controlled Clinical Trials. Wiley 2002

Pathways

Dinu V, PL Miller, H Zhao. Evidence for association between multiple complement pathway genes and AMD. Genetic Epidemiology 31: 224-237, 2007

Bar-Joseph Z, GK Gerber, TI Lee, NJ Rinaldi, JY Yoo, F Robert, DB Gordon, E Fraenkel, TS Jaakkola, RA Young, DK Gifford.Computational discovery of gene modules and regulatory networks. Nature Biotechnology 21:1337-1342, 2003

Carey VJ, J Gentry, E Whalen, R Gentleman. Network structures and algorithms in Bioconductor. Bioinformatics 21:135-36, 2005

Reed JL, I Famili, I Thiele, BO Palsson. Towards multidimensional genome annotation. Nature Reviews Genetics 7:130-141, 2006

Rodin A, TH Mosley, AG Clark, CF Sing, E Boerwinkle. Mining genetic epidemiology data with Bayesian networks application to *APOE g*ene variation and plasma lipid levels. Journal of Computational Biology 12:1-11, 2005

Stein CM, Y Song, RC Elston, G Jun, HK Tiwari, SK Iyengar. Structural equation model-based genome scan for the metabolic syndrome. *BMC Genetics* 2003, **4**(Suppl 1):S99

Sun N, H Zhao. Genomic approaches in dissecting complex biological pathways. Pharmacogenomics 5:163-79, 2004

Wang K, M Li, M Bucan: Pathway-Based Approaches for Analysis of Genomewide Association Studies. American Journal of Human Genetics 81:1278-1283, 2007

Zhao JH, J Luan, F Baksh, Q Tan. Mining gene networks with application to GAW15 problem 1. BMC Proceedings 1:S52, 2007

Study design

Brown BW, J Lovato, K Russell. Asymptotic power calculations: description, examples, computer code Statistics in Medicine 18:3137-51, 1999

Cai J, D Zeng. Sample Size/power Calculation for Case-cohort Studies. Biometrics 60:1015-1024, 2004

Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. American Journal of Epidemiology 155:478-484

Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. Statistics in Medicine 21:35-50, 2002

Gauderman WJ. Candidate gene association analysis for a quantitative trait, using parent-offspring trios. Genetic Epidemiology 25:327-338

Risch N, K Merikangas. The future of genetic studies of complex human diseases. Science 273:1516-1517, 1996

Self SG, RH Mauritsen, J Ohara. Power calculations for likelihood ratio tests in generalized linear models. Biometrics 48:31-39, 1992

R

Aulchenko YS, S Ripke, A Isaacs, CM van Duijn. GenABEL: an R library for genome-wide association analysis. Bioinformatics 23:1294-1296, 2007

Clayton D, HT Leung. An R package for analysis of whole-genome association studies. Human Heredity 64:45-51, 2007

Gonzalez JR, L Armengol, X Sole, F Guino, JM Mercader, X Estivill, V Moreno. SNPassoc: an R package to perform whole genome association studies. Bioinformatics, 23:644-645, 2007

Horton NJ, ER Brown ER, LJ Qian. Use of R as a toolbox for mathematical statistics exploration. American Statistician 58: 343-57, 2004

Ihaka R, R Gentleman. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics 5:299-314, 1996

Lake SL, H Lyon, K Tantisira, EK Silverman, ST Weiss, NM Laird, DJ Schaid. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. Human Heredity 55:56-65, 2003.

Warnes GR. The genetics package. R News 3: 9-13, 2003

Zhao JH. Use of R in genetics association studies. useR!2007, Ames, IA, USA

Zhao JH, Tan Q. Integrated analysis of genetic data with R. Human Genomics. 2:258-65, 2006

Zhao JH. Pedigree-drawing with R and graphviz. Bioinformatics 22:1013-1014, 2006

Zhao JH. Mixed-effects Cox models of alcohol dependence in extended families. BMC Genetics (Suppl) 6:127, 2005

Zhao JH. gap: genetic analysis package. J Statistical Software 23(8):1-18, 2007

SAGE

Elston RC, C Gray-McGuire. A review of the 'Statistical Analysis for Genetic Epidemiology' (S.A.G.E.) software package. Human Genomics 1:456-59, 2004

SAS

Chen Y-H, N Chatterjee, RJ Carroll. Retrospective analysis of haplotype-based case-control studies under a flexible model for gene-environment association. Biostatistics 9:81-99, 2008

Saxton AM. (Ed) Genetic Analysis of Complex Traits Using SAS. SAS Publications, 2004

Zhao JH, JA Luan, Q Tan, R Loos, N Wareham. Analysis of large genomic data in silico: The EPIC-Norfolk study of obesity. In DS Huang, L Heutte, and M Loog (Eds). Communication in Computer and Information Science 2, 781-790, 2007

Stata

Cleves MA. Exploratory analysis of single nucleotide polymorphism (SNP) for quantitative traits. Stata Journal 5:141-153, 2005

Feiveson AH. Power by simulation. Stata Journal 2:107-124, 2002

Harris RJ, MJ Bradburn, JJ Deeks, RM Harbord, DG Altman, JAC Sterne. Metan : fixed- and random-effects meta-analysis. Stata Journal 8(1) :3-28, 2008

Marchenko YV, RJ Carroll, DY Lin, CI Amos, RG Gutierrez. Semiparametric analysis of case–control genetic data in the presence of environmental factors. Stata Journal. 8:305-333, 2008

Mander AP. Haplotype analysis in population-based association studies. Stata Journal 1:58-75, 2001

Mander AP. Analysis of quantitative traits using regression and log-linear modeling when phase is unknown. Stata Journal 2:65–70, 2002

Saunders CL, DT Bishop, JH Barrett. Sample size calculations for main effects and interactions in case–control studies using Stata's nchi2 and npnchi2 functions. Stata Journal 3(1):47–56, 2003